Dense Scale Invariant Descriptors for Images and Surfaces

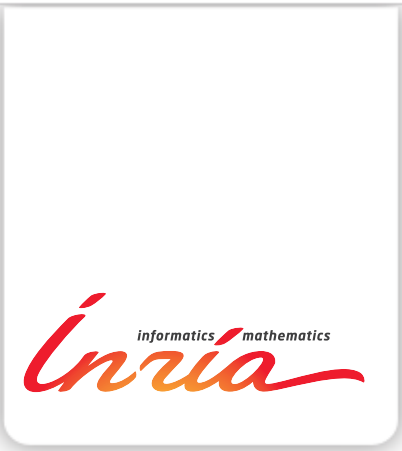
Iasonas Kokkinos, Michael Bronstein, Alan Yuille

Dense Scale Invariant Descriptors for Images and Surfaces

Iasonas Kokkinos, Michael M. Bronstein , Alan Yuille

RESEARCH REPORT
N° 7914
March 2012
Project-Teams GALEN

ISSN 0249-6399 ISRN INRIA/RR--7914--FR+ENG



Dense Scale Invariant Descriptors for Images and Surfaces

Iasonas Kokkinos*, Michael M. Bronstein† , Alan Yuille‡

Project-Teams GALEN

Research Report n° 7914 — March 2012 — 29 pages

Abstract: Local descriptors are ubiquitous in image and shape analysis, as they allow the compact and robust description of the local content of a signal (image or 3D shape). A common problem that emerges in the computation of local descriptors is the variability of the signal scale. The standard approach to cope with this is scale selection, which consists in estimating a characteristic scale around the few image or shape points where scale estimation can be performed reliably. However, it is often desired to have a scale-invariant descriptor that can be constructed *densely*, namely at every point of the image or 3D shape.
In this work, we construct scale-invariant signal descriptors by introducing a method that does not rely on scale selection; this allows us to apply our method at any point. Our method relies on a combination of logarithmic sampling with multi-scale signal processing that turns scaling in the original signal domain into a translation in a new domain. Scale invariance can then be guaranteed by computing the Fourier transform magnitude (FTM), which is unaffected by signal translations. We use our technique to construct scale- and rotation- invariant descriptors for images and scale- and isometry-invariant descriptors for 3D surfaces, and demonstrate that our descriptors outperform state-of-the-art descriptors on standard benchmarks.

Key-words: Scale invariance

* École Centrale Paris, France and INRIA-Saclay, France.
† Università della Svizzera Italiana, Lugano, Switzerland
‡ University of California at Los Angeles, USA and Korea University, Seoul, Korea

RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

Parc Orsay Université
4 rue Jacques Monod
91893 Orsay Cedex

Dense Scale Invariant Descriptors for Images and Surfaces

Résumé : Les descripteurs locaux sont omniprésents dans l'analyse d'image et de la forme, car ils permettent la description compacte et robuste du contenu local d'un signal (image ou une forme 3D). Un problème commun qui se dégage dans le calcul de descripteurs locaux est la variabilité de l'échelle du signal. L'approche standard pour faire face à cette probleme est la sélection d'échelle, qui consiste à estimer une échelle caractéristique autour des ces points d'image ou de la forme où l'estimation échelle peuvent être réalisées de manière fiable. Cependant, il est souvent souhaité d'avoir un descripteur invariant d'échelle qui peut être construit em densément, soit à chaque point de l'image ou la forme 3D.

Dans ce travail, nous construisons des descripteurs de signaux invariante d'échelle par l'introduction d'une méthode qui ne repose pas sur la sélection d'échelle; ce qui nous permet d'appliquer notre méthode à un point quelconque. Notre méthode repose sur une combinaison de l'échantillonnage logarithmique avec le traitement du signal multi-échelle qui transforme le changement d'échelle dans le domaine du signal original dans une translation dans un nouveau domaine. L'invariance d'échelle peut être garanti par le calcul de la magnitude de la transformée de Fourier (Fourier Transform Modulus -FTM), qui n'est pas affecté par les translations du signal.

Nous utilisons notre technique pour construire descripteurs invariantes de l'échelle et la rotation pour les images et les descripteurs invariantes de l'échelle et l'isométrie pour les surfaces 3D, et de démontrer que nos descripteurs peuvent surperformer l'état de l'art des descripteurs sur les benchmarks standards.

Mots-clés : Invariance d'échelle.

1 Introduction

Local descriptors have decisively pushed the envelope of solutions to computer vision and pattern recognition problems during the last decade. In image analysis, the emergence of descriptors that are robust to photometric changes and small displacements, able to cope with geometric (Euclidean or affine) transformations of the image, and the efficient coding schemes developed around them have facilitated the development of highly successful retrieval and matching applications [72, 24, 31, 59], while affecting the whole field of object recognition. In 3D shape analysis, 3D shape (surface) descriptors have been developed to describe local structure in a manner that is intrinsic, and thus invariant to isometric shape deformations, and used in applications such as shape retrieval [11].

A problem that emerges both in image and shape analysis when extracting descriptors is the variability in scale (signal resolution). Scale is a nuisance parameter, and it is commonly desirable to eliminate its effect on the descriptors. The standard approach for coping with scale has been *scale selection*, as in the seminal works of Lindeberg [40] and Lowe [44], where some low-level image processing criterion estimates a characteristic scale around a few salient interest points. This scale is then used to adapt the region over which the descriptor is computed, and thereby scale invariance is achieved. Scale selection was also used by Sun *et al.* [74] in their work on heat kernel signature (HKS) surface descriptors.

We argue that the scale selection strategy has many limitations both in image and shape analysis, since most structures in images and shapes do not lend themselves easily to reliable scale estimation, with the exception of some special cases. In image analysis, such cases are symmetric structures such as blobs or ridges where scale can be naturally correlated with the structure width. However, other structures such as edges are inherently 1D, so it is hard to estimate their scale. In practice, this means that when relying on a scale selection approach, scale-invariant local descriptors can be constructed only at a sparse set of points where scale can be estimated accurately. However, it is often desired to have a *dense* scale-invariant descriptor that can be constructed at every point of the image or 3D shape; moreover, regular sampling strategies have outperformed sparse (interest point-based) ones in an empirical evaluation [54].

In this paper, we take a different approach and compute scale-invariant descriptors without relying on scale selection. Instead we use a logarithmic (or log-polar) transformation of the signal domain that converts signal scalings (scalings and rotations, respectively) into translations, as shown in Figure 1. The effects of translations can then be eliminated using the properties of the Fourier transform.

We exploit this observation to build descriptors for two different applications. In the first part of the paper we present a dense scale- and rotation-invariant descriptor computed around image points. We build on the Daisy [77] descriptor to extract dense descriptors around all image points. Unlike Daisy, our descriptor is scale- and rotation- invariant; and unlike scale-adapted SIFT-type descriptors, our descriptor can be extracted around any image point. Our descriptor is thus both dense and invariant to image scalings and rotations, combining two desirable properties of SIFT and Daisy descriptors. A caveat related to our method is that our descriptor can get distorted around signal boundaries - we therefore assume that we are provided with large images and discard points on the image periphery. We demonstrate on a standard descriptor matching benchmark that our method outperforms state-of-the-art descriptors under a broad range of image transformations.

In the second part of the paper we introduce a scale-invariant descriptor for 3D surfaces, building on the intrinsic heat kernel signature (HKS) descriptor of [74]. By virtue of being intrinsic, i.e. depending solely on the Riemannian metric of the surface, HKS descriptors are invariant to isometric surface deformations, which include rotations and translations as special

cases. Moreover, HKS descriptors are applicable to a broad range of surface representations. We combine these favorable properties of HKS with scale invariance, giving rise to scale-invariant HKS (SI-HKS) [13]. By combining SI-HKS descriptors with a bag-of-words model [57], we obtain excellent performance on a non-rigid shape retrieval benchmark [8].

This paper presents the culmination of our early results in [36] and [13]. Comparing to [36], we present different low-level image measurements (from the monogenic signal [22] to a more common Daisy-based [77] front end), and extend the method to compute dense descriptors. Comparing to [13], we also study the use of volumetric intrinsic descriptors, and provide new results on the SHREC'10 benchmark [8].

We start our paper with an overview of existing works on local descriptors for images and surfaces in Section 2. We continue with a concise presentation of the scale-free scale-invariance principle for the case of a one-dimensional signal in Section 3 and proceed to describe how we use this idea to compute image and surface descriptors in Sections 4 and 5, respectively. In Section 6 we evaluate our approach on image and shape benchmark datasets.

2 Prior work

The main purpose of local descriptors in image and shape analysis is to summarize the information contained in the neighborhood of a point on a signal (image or 3D shape) into some low-dimensional feature vector, designed to be invariant to a certain class of transformations. For instance, image descriptors employ image derivatives and normalization to deal with additive and multiplicative illumination variations, respectively, while surface descriptors deal with isometric surface deformations by relying on intrinsic geometric quantities such as heat kernels. Furthermore, local descriptors can often deal graciously with occlusions, or boundary effects, where global descriptors, such as high-dimensional moments, fail. At the same time, it is substantially more challenging to deal with *transformations of the signal domain* such as scaling, or more generally, affine transformations.

In what follows we concisely describe the state-of-the-art in local image and 3D shape descriptors, and point out their shortcomings that our work aims at addressing.

2.1 Image Descriptors

Two seminal works in the development of image descriptors have been the Scale-Invariant Feature Transform (SIFT) [44] and Shape Contexts [68]. These descriptors perform 'soft' perceptual grouping, aggregating boundary information into a statistical description amenable to subsequent tasks, such as shape matching and object recognition.

These works have been refined and extended in multiple ways. Geometric Blur descriptors [5] extended Shape contexts to work with image gradients instead of shape contours. GLOH descriptors [51] used a log-polar grid to construct SIFT-like descriptors, SURF descriptors [3] used integral images for efficiency, while [80] used a search algorithm for the optimal combination of feature extraction, post-processing, and pooling. Moreover, dimensionality reduction [35], as well as linear [29, 15, 14] and nonlinear [60, 73] metric learning techniques have led to performance improvements, while using low-dimensional descriptors.

An important recent development includes the construction of *dense descriptors* over the whole image domain instead of single points. The computational complexity is handled using efficient convolution operations, both for rectangular [25] and for log-polar grids [77]. We note that these methods do not guarantee scale-invariance, since the descriptor scale is fixed beforehand. Our method addresses this shortcoming.

Coming to scale invariance, a two-stage process is commonly used, following the seminal work of [44]. First, a front end system is used to estimate local image scale, e.g. by using a scale-adapted differential operator [40]. Then, the estimated scale is used to adapt the domain over which the descriptor is computed. Variants of this idea include the identification of stable regions [47], entropy maximizing regions [32] or scale-invariant corners using the second moment-matrix [50]. In all cases a front-end process picks a certain scale which is then used for descriptor construction.

As we have argued in Section 1, this two-stage approach can often be problematic. In particular, around edges (and more generally, around non-symmetric points), there is no low-level criterion to reliably estimate image scale. Some notable exceptions include the works of [42, 52, 20], which, however, either implicitly exploit symmetry in order to work on edges ([52]), or require time-demanding operations to achieve invariance ([20, 42]). Instead, our approach provides a generic, feature-independent process to estimate scale-invariant descriptors, does not rely on ad-hoc scale-selection criteria, and can be efficiently implemented using the Fast Fourier Transform.

2.2 Shape Descriptors

The success of feature-based methods in image analysis has driven a more recent trend of developing similar methods for the analysis of 3D shapes in the computer graphics and geometry processing communities. Feature descriptors play an important role in shape correspondence and matching [76, 30] and retrieval [53, 57, 78], where the *bag of features* paradigm [72, 17] has been successfully employed.

On the one hand, it is possible to apply almost straightforwardly some successful image analysis methods to 3D shapes. Notable examples of feature detectors and descriptors that follow analogous methods in image analysis include corners [71] and edges [37], histograms of gradients [82] (akin to the use of [44] for images), 3D integral invariants [27] (first proposed in images in [46]), and maximally stable extremal regions (MSER) [19, 43].

On the other hand, it is possible to use 3D shape specific geometric structures for designing feature detectors and descriptors that have no direct image analogs. An ideal feature descriptor should be invariant to shape embedding in the 3D space, which includes *Euclidean invariance*, namely invariance to rotation and translation, and *deformation invariance*, namely invariance to inelastic deformations that preserve the metric structure of the surface (isometries). As deformation invariance guarantees Euclidean invariance, we focus on the former. Second, a descriptor should cope with missing parts, and also be insensitive to topological noise and connectivity changes in the shape (referred together as *topological invariance*). Third, it should work across different shape representations and formats (e.g. point clouds and meshes) and be insensitive to sampling (*representation invariance*). Finally, the descriptor should be invariant to global and local shape scaling (*scale invariance*).

Rotation and translation invariance can be achieved using volume and area descriptors [83], spherical harmonics [34], geometric moments [75], and distribution of pair-wise Euclidean distances [56].

Deformation invariance is more challenging. In [21], and later [48, 9] shapes were modeled as metric spaces with geodesic distances, which are invariant to inelastic deformations. This framework was used in [42] and [7] with a metric defined by internal distances in 2D shapes, while [66, 65] used the Laplacian spectra as intrinsic shape descriptors.

In [18] the authors popularize the notion of *diffusion geometry*, arising from the analysis of heat diffusion processes on manifolds and giving rise to intrinsic local and global structures. In [67, 45, 10, 12] global shape representations using different diffusion distances were proposed.

In [4] a conformal factor scalar descriptor is used, which is scale-invariant, but applicable

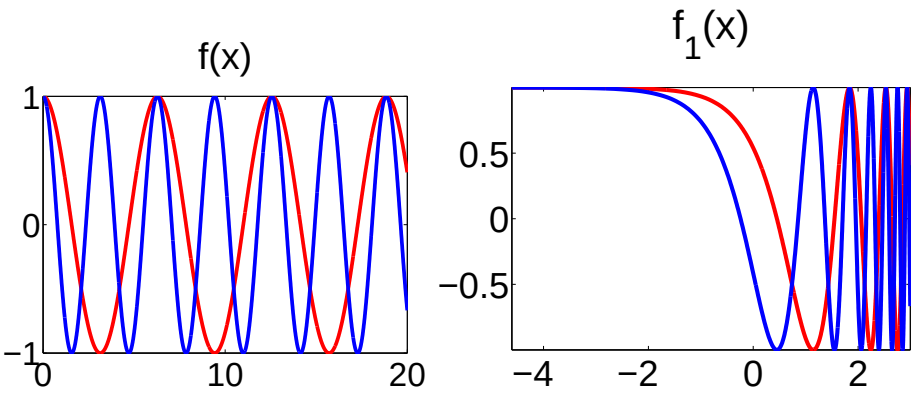


Figure 1: Demonstration of how the logarithmic transformation turns scalings into translations: on the left we have two functions $f(x) = \cos(x)$ (red),$g(x) = \cos(2x)$ (blue), differing by a scaling factor of two. By employing the signal transformation $f_a(x) = f(\log(ax))$ we obtain $f_1(x) = \cos(\log(x)), g_1(x) = \cos(\log(x) - \log 2)$, which only differ by a translation.

to shapes of fixed (e.g. sphere-like) topology and thus not topology-invariant. In [74] (and independently, [26]), local multiscale *heat kernel signature* (HKS) descriptors were constructed using the heat diffusion equation. This approach was later generalized to shapes with texture in [38]. HKS descriptors satisfy all of the above desired properties with the exception of scale invariance. In [2] the Shrödinger equation instead of the heat equation is considered, resulting in wave kernel signatures (WKS). Finally, [1, 6] show that both HKS and WKS can be regarded as a "filter" applied to the Laplace-Beltrami eigenvalues, and show that an optimal filter can be designed from examples by means of supervised learning.

In this paper, we show how to build a scale-invariant HKS descriptor. Our approach can be extended to other descriptors based on the Laplace-Beltrami operator.

3 Scale-Free Scale Invariance

Consider that we want to describe a one-dimensional signal $f(x),\ x > 0$ in a manner that does not change when the signal is scaled as $f(x/a),\ a > 0$. For this we consider a logarithmic transformation $h(x) = \log(x)$ of the domain. Denoting by f_a the transformation of $f(x/a)$, i.e. $f_a(h(x)) \doteq f(x/a)$, we have:

$$f_a(h(ax)) = f(x) = f_1(h(x)) \tag{1}$$

Since $a > 0, x > 0$ we can rewrite (1) as:

$$f_a(x' + \log(a)) = f_1(x'), \tag{2}$$

where $x' = \log(x)$. As illustrated in Figure 1, scaling the signal $f(x)$ by a thus amounts to translating the signal $f_1(x')$ by $-\log(a)$. We can then obtain a scale-invariant description of $f(x)$ in terms of the Fourier transform of $f_1(x')$; applying the shifting-in-time property of the Fourier transform to (2) gives:

$$\mathcal{F}_a(\omega)e^{j\log(a)\omega} = \mathcal{F}_1(\omega), \tag{3}$$

$$|\mathcal{F}_a(\omega)| = |\mathcal{F}_1(\omega)|, \tag{4}$$

where $\mathcal{F}_a(\omega)$ is the Fourier transform of $f_a(x)$. As per (4), the dependency on a has vanished.

This idea underlies the Fourier Transform Magnitude (FTM) technique, which has been used to deal with *global* image rotation and scaling in the context of image registration [16, 81] and texture classification [62]. One of our main contributions lies in applying this idea *locally*, at the level of local feature descriptors.

3.1 Discrete descriptors

So far our treatment has considered continuous signals. To deal with discrete signals we need to take into account the effects of sampling.

To avoid aliasing, the signals must be smoothed prior to sampling [55]. For equispaced samples the proper amount of smoothing is determined by the Nyquist theorem: the smoothing acts as a low-pass filter that cuts off the signal spectral content above the Nyquist frequency. However, in our case we sample the signal irregularly, namely at a set of locations forming a geometric progression, as detailed below in (6). We show that in our setting we can guarantee scale-invariance if we use a multi-scale smoothing scheme, where points close to 0 are subjected to small amount of smoothing, and points further out to stronger smoothing; this relates to the 'foveal' scale space of [41] as well as to Geometric Blur [5].

We consider the scale-space $\boldsymbol{f}(x, s)$ formed by convolving $f(x)$ with a set of kernels $g_s(x) = \frac{1}{s} g_1(x/s)$:

$$\boldsymbol{f}(x, s) = f(x) * g_s(x). \tag{5}$$

We propose to sample $f(x)$ through $\boldsymbol{f}(x, s)$ as follows:

$$f[n] = \boldsymbol{f}(c_0 \alpha^n, c\alpha^n), \tag{6}$$

where $f[n]$ is the sampled version of our original signal $f(x)$. As mentioned above, we use a geometric spacing of samples while smoothing proportionally to the distance from 0.

Denoting by $f_\alpha[n]$ the signal obtained by sampling $f(x/\alpha)$, we prove in the Appendix that $f_{\alpha^k}[n] = f_1[n-k], \forall n, k$. This allows us to adapt the technique described in the previous subsection and use the Discrete-Time Fourier Transform (DTFT) [55] of $f_\alpha[n]$ and $f_1[n]$ instead of the Continuous-Time Fourier Transform discussed above.

Finally, since we work with finitely supported signals, we face signal boundary effects; we discuss these as appropriated to image and surface descriptors.

4 Scale-Invariant Image Descriptors

In Section 3 we have described a general approach to achieving scale-invariance for 1D signals. We now elaborate on how we exploit this to achieve scale- and rotation- invariance on images.

For this, we construct a descriptor around a point by sampling its neighborhood with a log-polar grid, as shown in Figure 2. This sampling scheme turns image rotations/scalings into translations, allowing us to use the FTM technique to achieve rotation- and scale- invariance. We note that the log-polar transform has already been used to deal with *global* scale and rotation changes [16, 62, 81, 33], while in [69] it is associated with the foveal sampling pattern of the retina. Our contribution lies in exploiting this scheme for *local* descriptor construction.

As illustrated in Figure 2, we construct a descriptor around a point $\mathbf{x} = (x_1, x_2)$ by sampling its neighborhood along K rays leaving $\mathbf{x}$ at equal angle increments $\theta_k = 2\pi k/K$. Along each ray we sample the image at N points with distances $r_n = c_0 a^n$ from $\mathbf{x}$, using distance-dependent

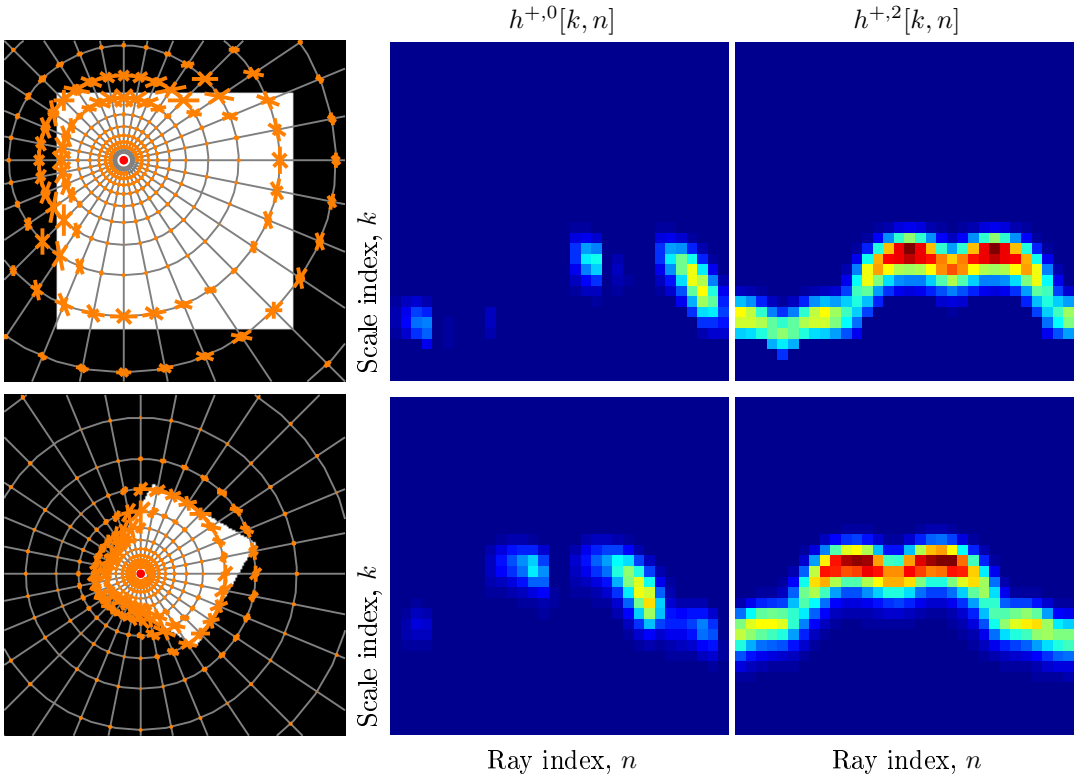


Figure 2: Effect of scale and orientation changes on log-polar descriptors. On the left we show as needle diagrams the descriptors computed on a point before and after scaling and rotating an image; needle length is proportional to the quantity in (9). In the next two columns we show two of the descriptor components computed by (9). The effect of scaling and rotation amounts to a translation, which can be eliminated by using the 2D Fourier transform. As the point is arbitrary (i.e. not a corner/junction/blob center), performing scale selection around it would be far from obvious.

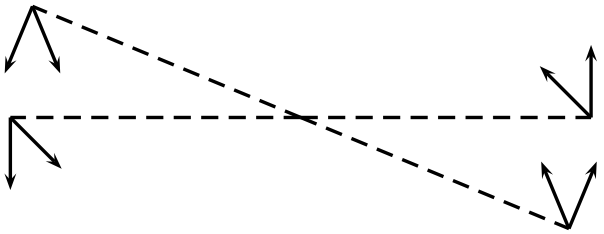


Figure 3: The orientations used in directional derivative computation are set relative to ray orientation, so that image rotation will amount to feature translation; here we show the first two directions on four different rays.

smoothing with a 2D Gaussian kernel of standard deviation $\sigma_n = \alpha r_n$. These measurements form a $K \times N$ matrix:

$$h[k,n] = f_{\alpha r_n}\left[x_1 + r_n \cos(\theta_k), x_2 + r_n \sin(\theta_k)\right], \tag{7}$$

where f_σ denotes the sampled field f smoothed by σ.

Since we sample according to Section 3.1, rescaling the signal f will amount to shifting h along the second (n/scale) dimension. Moreover, since the rays are equally spaced, rotating the image around $\mathbf{x}$ by $\frac{2\pi}{K}$ amounts to shifting h along the first (k/orientation) dimension by one. Image scalings and rotations thus turn into horizontal and vertical translations of h.

We can therefore use the FTM technique to locally describe the image in a scale- and rotation-invariant manner. In specific, from the time-shifting property of the DTFT [55], we know that if $h[k,n] \overset{\mathcal{F}}{\leftrightarrow} H(j\omega_k, j\omega_n)$ are a DTFT pair, we will then have:

$$h[k-c, n-d] \overset{\mathcal{F}}{\leftrightarrow} H(j\omega_k, j\omega_n) e^{-j(\omega_k c + \omega_n d)} \tag{8}$$

so taking the absolute of the DTFT yields a scale- and rotation- invariant quantity.

In practice our descriptor is not perfectly invariant to scale changes, due to the limited number of scales considered. As the right columns of Figure 2 illustrate, scaling an image introduces new observations at fine scales and removes others at coarse scales. As experimentally demonstrated in Section 6, despite this approximation, our descriptor systematically outperforms SIFT for scale changes up to an order of 3.

Furthermore, descriptors lying close to the image boundaries can be largely affected by boundary effects; we restrain our evaluation to descriptors lying sufficiently far from the boundaries, leaving the proper imputation of missing values to future work.

4.1 Directional Derivative Information

The descriptor described so far is scale- and rotation- invariant, but not illumination invariant. Following [5, 14, 77], we combine the computation of directional derivatives with polarization, i.e. separating the negative and positive parts. Directional derivatives are invariant to additive intensity changes and provide information about the signal's dominant orientations. Polarization gave consistent improvements [14], which can be attributed to segregating upwards and downwards trends in the signal values.

Using directional derivatives has two implications. First, we need to align the directional derivatives with the ray directions as shown in Figure 3; this ensures that image rotation amounts

to feature translation - i.e. the features are rotation-covariant. Concretely, on ray k we use as d-th feature orientation the angle $\theta_{d,k} = k\frac{2\pi}{K} + d\frac{2\pi}{D}$. The computation of directional derivatives at all angles is efficiently performed by exploiting the steerability of directional derivatives of the Gaussian.

Second, the magnitude of the derivative signals will decrease for larger scales, as the derivated signal becomes increasingly smooth. An analysis similar to [40] therefore suggests scaling the output of smoothing by the root of the kernel's standard deviation, σ.

Coming to the treatment of multiplicative image changes, we normalize by dividing with the L_2 norm of the descriptor elements. As also mentioned in [77], we have observed that normalizing over each ring separately gives better results than normalizing over all rings.

4.2 Descriptor Construction

We first briefly summarize the steps used for the construction of our descriptor. We omit some technical implementation aspects, as we make our image descriptor code publicly available.

For a given scale n we smooth the image at a scale $\sigma_n = ar_n, \quad r_n = c_0 a^n$, giving rise to I_{σ_n}. To gather samples on ray k and scale n we compute the directional derivatives of I_{σ_n} for D orientations, $\theta_{d,k} = k\frac{2\pi}{K} + d\frac{2\pi}{D}$, scale them by σ_n and split them into positive and negative parts. This gives $2D$ signals, $f_{\sigma_n}^{\pm,\theta_{d,k}}$, where $\pm$ stands for polarity, $\theta_{d,k}$ indicates the derivative orientation, and σ_n the smoothing scale.

These signals are then used instead of f in (7), giving rise to $2D$ measurements per $[k,n]$ combination:

$$h_{\mathbf{x}}^{\pm,d}[k,n] = \sigma_n f_{\sigma_n}^{\pm,\theta_{d,k}}\left[x_1 + r_n\cos(\theta_k), x_2 + r_n\sin(\theta_k)\right], \tag{9}$$

where $k \in \{1,\dots K\}, n \in \{1,\dots N\}$, with K the number of rays and N the number of rings. Put together, this gives a $2KDN$-dimensional descriptor per point. We form our descriptor at $\mathbf{x}$ by (a) computing the 2D Discrete Fourier Transform (DFT) for every $[\pm, j]$ combination, (b) keeping the DFT elements corresponding to DTFT frequencies $(\omega_k, \omega_n) \in [0, 2\pi] \times [0, \pi)$ - due to the symmetry of the Fourier transform and (c) concatenating all components in a single vector.

We use $N = 36$ rays, $D = 4$ orientations and $K = 28$ rings, which results in a 7168-dimensional descriptor. This descriptor is largely redundant and could be compressed using dimensionality reduction techniques [29, 15, 60, 14, 73] - we leave this for future work.

Coming to computational cost, we note that the features required to form our descriptor in (9) can be obtained in batch mode with efficient recursive filtering [28] and steering. As an indicative measure of relative time performance, we report the timings for the demo script contained in our distribution: for an image of size 700x1000 all convolutions cost approximately 0.8 seconds; the formation of ~ 23000 ($136\times$ 170 regularly-spaced) descriptor elements 1.2 seconds; steering and normalization 6.2 seconds; and the FFT 3 seconds. In sum, we obtain 23000 scale-invariant descriptors in close to 10 seconds. The most time-consuming part is Matlab-based, and could be accelerated in C, while the descriptor construction is easy to parallelize.

In Figure 4 we show the values of the lowest frequency (and highest-energy) coefficients of densely computed descriptors, as evaluated on two images. We see that their values are effectively invariant to image rotations and scalings, despite a scaling factor in the order of 2. In the bottom row, we use the two points on the left image as references and 'query' the right one for points having similar descriptors. We then show the similarity of 'query' point descriptors to the red (left) and the green (right) reference points, which indicates the discriminative ability of the descriptor - even though locally the structures are similar, the context helps disambiguate them.

A thorough experimental evaluation of our descriptor follows in Section 6.1.

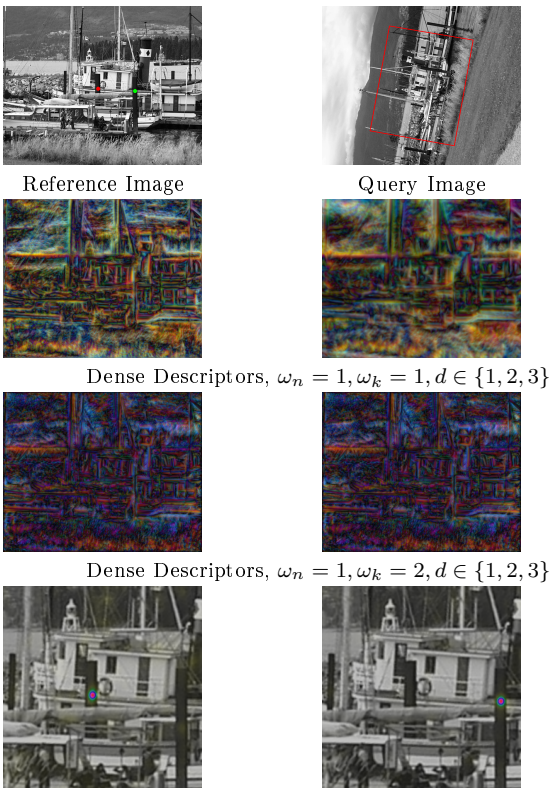


Figure 4: Visualization of dense scale- and rotation- invariant descriptors: we show some our descriptor dimensions as R, G, and B channels, and compare the resulting images over a reference image and the corresponding locations of a query image. In the bottom row we superimpose on the image hue maps (red and colorful is larger) indicating the similarity of descriptors in the query image to the points on the reference image. Based on context information, our descriptor can discriminate among locally similar structures.

5 Scale-Invariant Shape Descriptors

In this Section we detail how Scale-Free Scale-Invariance can be applied to the computation of scale-invariant local shape descriptors. As mentioned in Section 2, the Heat Kernel Signatures of [74] satisfy *deformation invariance* and *representation invariance*, but not scale invariance. Our contribution consists in making HKS scale-invariant. We first provide a concise description of HKS alongside with their numerical computation, and then describe our scale-invariant modification of HKS.

5.1 Heat Kernel Signatures

Let us model the shape of a 3D physical object as a connected and compact region $X \subset \mathbb{R}^3$, whose boundary ∂X is a closed connected two-dimensional Riemannian manifold. Traditionally, shape deformations are modeled as isometries of the 2D boundary surface ∂X preserving its Riemannian metric structure ∂X (we refer to such deformations as *boundary isometries*). Such

Figure 5: Invariance of the first three components of the HKS (shown as R, G, and B channels, respectively), for a shape undergoing isometric transformations.

deformations can bend the surface of the shape but not stretch it; however, the volume bounded by ∂X can change. Works on feature descriptors thus mainly focus on defining such geometric structure that would be intrinsic (i.e., expressible solely in terms of the Riemannian metric of ∂X) and consequently, invariant to boundary isometries of ∂X.

A recent line of works [66, 18, 39, 67, 58, 74, 13] studied intrinsic descriptions of shapes by analyzing heat diffusion processes on ∂X. Such processes give rise to the so-called *diffusion geometry* and arise from the *heat equation*

$$\left(\frac{\partial}{\partial t} + \Delta_{\partial X}\right) u(t,x) = 0, \tag{10}$$

where $u(t,x) : [0,\infty) \times \partial X \to [0,\infty]$ is the heat value at a point x in time t, and $\Delta_{\partial X}$ is the positive-semidefinite Laplace-Beltrami operator associated with the Riemannian metric of ∂X. The Laplace-Beltrami operator $\Delta_{\partial X}$ is *intrinsic* to the two-dimensional manifold ∂X, meaning that it is expressible in terms of the metric of ∂X. Consequently, it is invariant under boundary isometries.

The solution $h_t(x,y)$ of (10) corresponding to a point initial condition $u(0,x) = \delta(x,y)$, is called the *heat kernel* and represents the amount of heat transferred on ∂X from x to y in time t due to the diffusion process.

In particular, the diagonal of the heat kernel $h_t(x,x)$ (i.e., setting $x = y$, also referred to as the *auto-diffusivity*) describes the amount of heat remaining at point x after time t. Its value is related to the Gaussian curvature $K(x)$ through $h_t(x,x) \approx \frac{1}{4\pi t}\left(1 + \frac{1}{6}K(x)t + \mathcal{O}(t^2)\right)$, expressing the well-known property that heat tends to diffuse slower at points with positive curvature, and faster at points with negative curvature.

The spectral decomposition of the Laplace-Beltrami operator $\Delta_{\partial X}$ produces a set of orthonormal eigenfunctions $\phi_0 = \text{const}, \phi_1, \phi_2, ...$ and corresponding eigenvalues $\lambda_0 = 0 \leq \lambda_1 \leq \lambda_2 ...$ satisfying $\Delta_{\partial X}\phi_i = \lambda_i \phi_i$. These eigenfunctions form a basis on $L_2(\partial X)$ analogous to the Fourier

basis on Euclidean domains, where the heat kernel has the following expansion [39]:

$$h_t(x,y) = \sum_{i\geq 0} e^{-\lambda_i t}\phi_i(x)\phi_i(y). \tag{11}$$

Sun *et al.* [74] used the diagonal of the heat kernel as a local surface descriptor referred to as *heat kernel signature* (HKS). The HKS descriptor at each point x of the surface ∂X is defined as a q-dimensional vector

$$H(x) = (h_{t_1}(x,x), ..., h_{t_q}(x,x)), \tag{12}$$

where $t_1, ..., t_q$ is a set of time scales. As illustrated in Figure 5, $H(x)$ captures multi-scale shape curvature information in a isometry-invariant manner. Furthermore, by virtue of being intrinsic, the HKS is invariant to boundary isometries of ∂X. Equation (11) allows for efficient computation of the heat kernel, which in practice requires computing the first few eigenpairs of the Laplace-Beltrami operator.

5.2 Volumetric Heat Kernel Signatures

Raviv *et al.* [64] argued that a smaller class of *volume isometries* preserving the metric structure inside the volume X are more suitable for modeling realistic shape deformations that boundary isometries, which preserve the area of ∂X, but not necessarily the volume of X (volume isometries are necessarily boundary isometries, but not vice versa). Thus, instead of considering diffusion processes on the boundary surface ∂X, diffusion *inside* the volume X, arising from the Euclidean volumetric heat equation with Neumann boundary conditions,

$$\begin{aligned} \left(\frac{\partial}{\partial t} + \Delta\right) U(t,x) &= 0 \qquad && x \in \mathrm{int}(X); \\ \langle \nabla U(t,x), n(x)\rangle &= 0 && x \in \partial X, \end{aligned} \tag{13}$$

was considered in [64] (here, $U(t,x) : [0,\infty) \times \mathbb{R}^3 \to [0,\infty]$ is the volumetric heat distribution, Δ is the Euclidean positive-semidefinite Laplacian, and $n(x)$ is the normal to the surface ∂X at point x. In the following, we use capital letter to denote the volumetric quantities). The heat kernel of the volumetric heat equation (13) is given, similarly to (11) by

$$H_t(x,y) = \sum_{i\geq 0} e^{-\Lambda_i t}\Phi_i(x)\Phi_i(y), \tag{14}$$

where Φ_i and Λ_i are the eigenfunctions and eigenvalues of Δ satisfying $\Delta\Phi_i = \Lambda_i\phi_i$ and the boundary conditions $\langle \nabla\Phi_i(x), n(x)\rangle = 0$. The diagonal of the heat kernel $H_t(x,x)$ gives rise to the *volumetric HKS* (vHKS) descriptor [64], which is invariant to volume isometries of X. Compared to the 2D HKS, such descriptors were shown to be less sensitive to geometric and topological noise [64].

5.3 Scale invariance

A notable disadvantage of heat kernel descriptors (both HKS and vHKS) is their sensitivity to scale. Given a shape X and its version X' uniformly scaled by the factor of a, it is easy to establish[1] that the eigenfunctions and eigenvalues are scaled inversely proportionally to the area

[1] The first relation stems from normalization of the Laplace-Beltrami operator; the second relation stems from the unit norm of the eigenfunctions, $\|\phi_i\|_{L_2(\partial X)} = \|\Phi_i\|_{L_2(X)} = 1$.

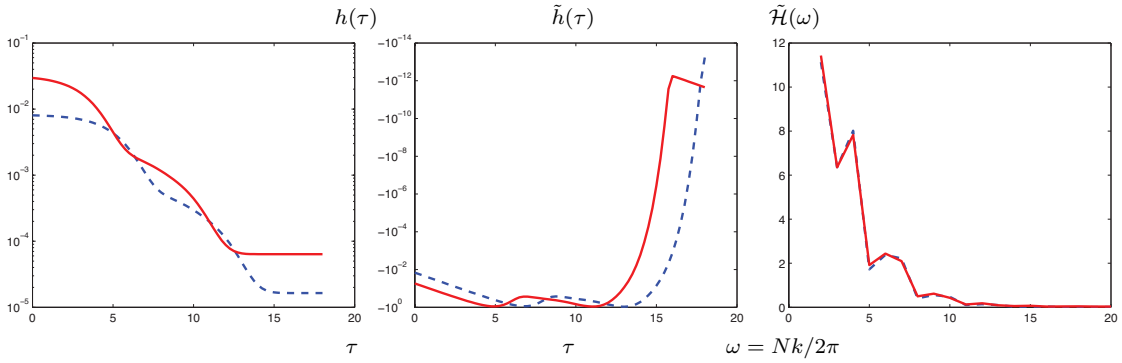


Figure 6: Construction of the SI-HKS. Left: HKS computed at the same point, for a shape that is scaled by a factor of 11. Middle: the signal $\dot{h}[\tau]$, where the effects of scale change have been converted into a shifting in time. Right: the first 20 components of $|H(\Omega)|$ for the two signals; the two descriptors are virtually identical.

of ∂X ($\propto a^2$) or the volume of X ($\propto a^3$), respectively. Thus, the new set of the eigenfunctions and eigenvalues of X' is given by

$$\lambda' = a^{-2}\lambda; \;\; \phi' = a^{-1}\phi; \tag{15}$$

$$\Lambda' = a^{-2}\Lambda; \;\; \Phi' = a^{-3/2}\Phi, \tag{16}$$

so the corresponding heat kernels satisfy

$$h'_t(x) = \sum_{i=0}^{\infty} e^{-\lambda_i a^{-2} t}\phi_i^2(x)a^{-2} = a^{-2}h_{a^{-2}t}(x), \tag{17}$$

$$H'_t(x) = \sum_{i=0}^{\infty} e^{-\Lambda_i a^{-2} t}\Phi_i^2(x)a^{-3} = a^{-3}H_{a^{-2}t}(x), \tag{18}$$

relating the signature h' (respectively, H') at time t for X' with the signature h (respectively, H) at time $a^{-2}t$ for X.

Typically, the scaling factor a is unknown *a priori*. Scale dependence can be removed by *global* normalization, for example, dividing the eigenfunctions and eigenvalues by the first non-zero eigenvalues, $\lambda_i = \lambda_i\lambda_1^{-1}$ and $\phi_i = \phi_i\lambda_1^{-1}$ ($\Lambda_i = \Lambda_i\Lambda_1^{-1}$ and $\Phi_i = \Phi_i\Lambda_1^{-1}$, respectively). However, global normalization does not work in cases when the shape undergoes transformations changing its global geometry, such as removing significant parts – in these cases, the resulting eigenvalues and eigenfunctions can be very different. One thus has to resort to local normalization.

One possibility is to locally normalize the metric structure of the manifold. This approach has been explored by Raviv *et al.* [63], who constructed an affine-invariant Riemannian metric tensor on the manifold (rather then using the metric induced by the embedding) and derived an affine-invariant diffusion geometry form the associated Laplace-Beltrami operator. Another possibility is, similarly to image descriptors, to estimate scale locally from a feature detection algorithm. However, in shape analysis there is no clear analogue for features such as blobs or corners, and such an estimation is not always possible. Moreover, as mentioned before, in many cases we are interested in dense descriptors computed at all points.

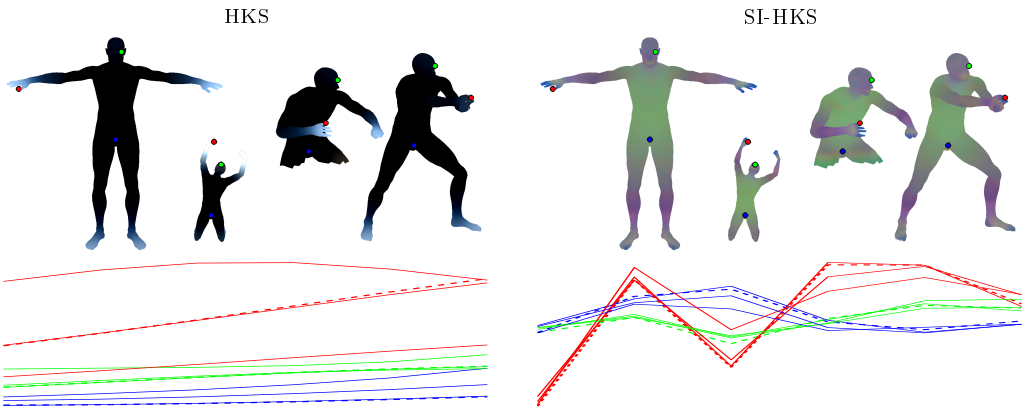


Figure 7: Top: three components of the HKS (left) and the proposed SI-HKS (right), represented as RGB color and shown for different shape transformations (null, isometric deformation+scale, missing part, topological transformation). Bottom: HKS (left) and SI-HKS (right) descriptors at three points of the four shapes (different points are coded with red, green, and blue; dashed line shows the null shape descriptor). We observe that the SI-HKS descriptors are substantially more robust to the deformations and stay closer to the null shape descriptor.

5.4 Scale-invariant Heat Kernel Signatures

We now describe a method to discard the dependence of h (or H) on the unknown scaling factor a following the approach outlined in Section II, resulting in a *scale-invariant heat kernel signature* (SI-HKS). We develop our construction for the HKS; scale-invariant vHKS can be obtained in a similar way.

First, at each point x we sample the heat kernel scale logarithmically with some basis α, denoted here as $h(\tau) = h_{\alpha^\tau}(x,x)$. In this scale-space, the heat kernel of the scaled shape becomes $h'(\tau) = a^{-2}h(\tau + 2\log_\alpha a)$ (Fig. 6, left). Second, in order to remove the dependence on the multiplicative constant a^{-2}, we take the logarithm of the signal and then differentiate it w.r.t. the scale variable,

$$\begin{aligned}\frac{d}{d\tau}\log h'(\tau) &= \frac{d}{d\tau}\left(-2\log a + \log h(\tau + 2\log_\alpha a)\right)\\ &= \frac{\frac{d}{d\tau}h(\tau + 2\log_\alpha a)}{h(\tau + 2\log_\alpha a)}.\end{aligned} \tag{19}$$

Denoting

$$\tilde{h}(\tau) = \frac{\frac{d}{d\tau}h(\tau)}{h(\tau)} = \frac{-\sum_{i\geq 0}\lambda_i\alpha^\tau\log\alpha e^{-\lambda_i\alpha^\tau}\phi_i^2(x)}{\sum_{i\geq 0}e^{-\lambda_i\alpha^\tau}\phi_i^2(x)},$$

we thus have a new function $\tilde{h}$ which transforms as $\tilde{h}'(\tau) = \tilde{h}(\tau + 2\log_\alpha a)$ as a result of scaling (Fig. 6, center). Finally, using the idea of Section II, we apply the Fourier transform to $\tilde{h}$ and take its absolute value,

$$\mathcal{F}[\tilde{h}'](\omega) = \tilde{\mathcal{H}}'(\omega) = \tilde{\mathcal{H}}(\omega)e^{-j\omega 2\log_\alpha a}, \tag{20}$$

$$|\tilde{\mathcal{H}}'(\omega)| = |\tilde{\mathcal{H}}(\omega)|, \tag{21}$$

producing a scale-invariant descriptor (Fig. 6, right).

One caveat of this approach could be that scaling the shape and then resampling the function $\dot{h}[\tau]$ makes the samples at the boundaries change. Fortunately, the HKS is smooth at low- and high- scales and therefore its derivative is equal to zero for a broad range of τs at the beginning and end of $\dot{h}$. In Fig. 6 we show the intermediate signals involved in the construction of the SI-HKS.

5.5 Numerical computation

Due to the possibility to express the heat kernel in the spectral domain, the practical computation of HKS or vHKS and their scale-invariant versions boils down to discretizing the Laplacian of the shape, computing its first eigenvectors and eigenvalues, and approximating formulae (11) or (14) using a finite number of terms (since the exponential coefficients, only a small number of eigenfunctions and eigenvalues is required). Since there exists a plethora of methods for Laplacian discretization on different representations of shapes (in particular, meshes, point clouds, volumes, and implicit surfaces), the heat kernel descriptors are very versatile and, up to errors of the particular Laplacian approximation, representation-invariant.

In the case of surfaces represented as point clouds $V = \{v_1, ..., v_N\} \subset \partial X$ or triangular meshes, the discretization of the Laplace-Beltrami operator of the surface ∂X can be written in the generic matrix-vector form as $\Delta_{\partial X} f = A^{-1} W f$, where $f = f(v_i)$ is a vector of values of a scalar function $f : \partial X \to \mathbb{R}$ sampled on the vertices, $W = \text{diag}\left(\sum_{l \neq i} w_{il}\right) - (w_{ij})$ is a zero-mean $N \times N$ matrix of weights, and $A = \text{diag}(a_i)$ is a diagonal matrix of normalization coefficients [23, 79]. A particular choice that is popular in computer graphics for triangular meshes is the *cotangent weight* scheme [61, 49], where

$$w_{ij} = \begin{cases} (\cot \alpha_{ij} + \cot \beta_{ij})/2 & (v_i, v_j) \text{ is an edge;} \\ 0 & \text{else,} \end{cases} \tag{22}$$

where α_{ij} and β_{ij} are the two angles opposite to the edge between vertices v_i and v_j in the two triangles sharing the edge, and a_i are the discrete area elements. The eigenfunctions and eigenvalues of $\Delta_{\partial X}$ are found by solving the generalized eigendecomposition problem $W\phi_i = A\phi_i\lambda_i$ [39].

In the volumetric case, the shapes are rasterized and represented as arrays of voxels on a regular Cartesian grid, allowing to use the standard Euclidean Laplacian. In [64], the volumetric Laplacian was discretized using a 6-neighborhood stencil, and boundary conditions were enforced using the shadow variables technique.

6 Results

6.1 Image Descriptor Evaluation

We use the dataset, code and protocol of [51] to evaluate descriptor performance: ground truth correspondences between two images of an identical scene are used to evaluate interest point matches found based on descriptor similarities. Even though our descriptors can be computed densely, we use the Hessian- and Harris- Laplace interest point operators of [51], in order to compare with SIFT descriptors on equal grounds. We do not compare with Hessian- and Harris-affine detectors, as our descriptors are not designed to cope with affine transformations.

An issue regarding evaluation is that these two detectors provide as output points associated with scale ('regions'), while our descriptor is only using point information in its construction. Two points lying at the same location (modulo the transformation registering the images), but

	$d_{i,j} < \theta_d$ (Ours ✓)	$d_{i,j} \geq \theta_d$ (Ours ×)
$O_{i,j} \geq \theta_0$ (SIFT ✓)	✓	×
$O_{i,j} < \theta_0$ (SIFT ×)	×	✓

Table 1: Conditions for incorporating pair i, j in the evaluation. $d_{i,j}$ stands for the Euclidean distance between the centers of points i, j and $O_{i,j}$ for the overlap of their regions, while θ_d and θ_O are the respective thresholds.

with very different scales are considered different by the evaluation approach of [51]: a match of two such points is declared a false positive. However, as far as our descriptors can tell, these two such points are identical, as they correspond to the same scene point. On the flip side, due to the log-polar sampling, our descriptors are location-sensitive, while SIFT descriptors of large regions are less affected by displacements.

To make the two methods of computing scale-invariant descriptors commensurate we distinguish four cases, as indicated in Table 1. In our evaluation we only consider pairs of points where either both criteria are met (point centers are close, and their areas overlap substantially), or both criteria are violated (points centers are far, and their areas have low overlap). We exclude points where one criterion is met, while another is not (e.g. having close centers, but low area overlap). This amounts to some 5-10% of positive point pairs being excluded from the evaluation.

Moreover, as mentioned in Section 4.2, our descriptor gets distorted around the image boundaries. One obvious remedy could be to decrease the size of the used log-polar grid. This comes at the cost of reducing the contextual information captured by our descriptor, while limiting the range of scale changes that can be dealt with.

In our experiments we use images of approximately 700x1000 and 1000x1000 pixels, while our descriptor's maximal ring size is at a radius of 230 pixels. In our evaluation we only consider descriptors of points which lie at least 100 pixels from the closest image boundary, accounting for roughly 60%-85% of points within an image. This ensures that the associated descriptors are not largely distorted, and deconvolves the evaluation of our descriptors scale-invariance from the boundary effects. When cutting the pixel distance from 100 down to 50 we observed a deterioration of our results, in particular for large scale changes.

Three alternative methods are proposed in [51] to find correspondences. The simplest one ('similarity') computes all distances between the descriptors in the two images and declares as matches all pairs of points whose distance is below a given threshold. A more elaborate technique ('k-nearest') first solves a linear assignment problem, allowing a descriptor in one image to match at most with a single descriptor in the other. For a given value of k, the k best-matching descriptors are computed according to this procedure. Finally the most elaborate technique ('distance ratio') normalizes the distances between a descriptor and descriptors in another image by the distance of the descriptor to its nearest-neighbor in the other image. We report results using all three criteria, in the form of Precision-Recall curves; we obtain these curves from the software of [51], adapted to include the evaluation modifications described above.

In Figure 8 we first examine the robustness of our descriptors to transformations other than scaling and rotation, including blurring due to a change of camera focus, jpeg compression, and perspective transformations. We observe that according to all three criteria, our descriptor outperforms SIFT.

To assess the robustness of our descriptor to scale and rotation changes we use all of the images in [50] and subject them to a common set of synthetic transformations. In the original dataset only two out of eight images were subjected to scaling and rotation, and the amounts of scaling were not common across the two images; this hindered the thorough evaluation of our

method, and led us to use this more controllable setting instead.

In Figure 10 we demonstrate how our descriptor compares to SIFT, for increasingly large changes in scale. We observe that our descriptor largely outperforms SIFT according to all three criteria, up to a scale change in the order of 3 (green color). Above that level, the results become ambiguous, with SIFT performing equally well or better on most images.

In summary, the results verify that our descriptor has excellent performance for a broad range of scales -at least up to three-, despite being only approximately scale invariant (due to the use of a limited number of rings/scales). This complements the main merit of our approach, namely than unlike SIFT, or other descriptors which rely on scale selection, our descriptor can be evaluated anywhere over the image domain.

6.2 Shape Descriptor Evaluation

We used the SHREC 2010 robust large-scale shape retrieval benchmark, simulating a retrieval scenario, in which the queries include multiple modifications and transformations of the same shape [8]. The shapes were represented as triangular meshes with the number of vertices ranging approximately between 300 and 30,000. The dataset consisted of two parts: 715 shapes from 13 shape classes with simulated transformation (55 per shape) used as queries and the remaining 456 shapes. The query set consisted of 13 shapes taken from the dataset (null shapes), with simulated transformations of different type and strength applied to them. Each query had only one correct corresponding null shape in the dataset.

Performance was evaluated using mean average precision and the receiver operating characteristic (ROC). *Precision* $P(r)$ is defined as the percentage of relevant shapes in the first r top-ranked retrieved shapes. In the present benchmark, a single relevant shape existed in the database for each query. *Mean average precision* (mAP) is defined as $mAP = \sum_r P(r) \cdot rel(r)$, ($rel(r)$ is the relevance of a given rank), and ideally should be 100%. The *receiver operating characteristic* (ROC) curve is another performance criterion, representing a tradeoff between the percentages of similar shapes correctly identified as similar (*true positives rate* - TPR) and of dissimilar shapes wrongfully identified as similar (*false positive rate* - FPR).

Heat kernel signatures (HKS) and the proposed scale-invariant heat kernel signatures (SI-HKS), respectively, were used as local shape descriptors. In both cases, the cotangent weight scheme was used to discretize the surface Laplace-Beltrami operator $\Delta_{\partial X}$. The heat kernel was approximated using the $k = 100$ largest eigenvalues and eigenvectors. For HKS, we used the parameters of [57] (six scales $1024, 1351, 1783, 2353, 3104$ and 4096), which were experimentally found to give optimal performance. We construct the SI-HKS as described in Section 5, using a logarithmic base $\alpha = 2$ and τ ranging from 1 to 25 with increments of $1/16$. The first 6 lowest frequencies of the Fourier transforms were used.

Bag-of-features shape descriptors were constructed using bags of geometric words proposed in [57]. For HKS and SI-HKS, a geometric vocabulary of size 48 was built using clustering in the signature space (six-dimensional in both cases). The HKS and SI-HKS at each point of the shape were replaced by the closest geometric word from the vocabulary using soft vector quantization. The distribution of geometric words (48-dimensional bag of features) was used as the shape descriptor. The L_1 distance was used to compare the bags of features.

Tables 2–3 show the performance of shape retrieval using bags of features based on HKS and SI-HKS local descriptors. SI-HKS shows dramatic improvement (from 27.42% to 98.21% MAP and from 30.34% to 65.07%) in the *scale* and *mixed* transformations classes, respectively, and a small improvement (from 80.22% to 82.08% and from 2.95% to 6.61%) in the *local scale* and *partial* classes, respectively. An insignificant performance degradation is manifested in *topology, holes*, and *sampling*. Overall in all transformation classes and strengths in the SHREC benchmark,

SI-HKS performs better than HKS (90% vs 85.00%). These results are consistent with the ROC curves shown in Figure 12.

Table 2: Performance (mAP in %) of ShapeGoogle using bags of features of size 48 based on HKS local descriptor.

	Strength				
Transform.	**1**	**≤2**	**≤3**	**≤4**	**≤5**
Isometry	100.00	100.00	100.00	100.00	100.00
Topology	100.00	98.08	97.44	96.79	96.41
Holes	100.00	100.00	97.44	95.19	90.13
Micro holes	100.00	100.00	100.00	100.00	100.00
Scale	0.98	40.68	43.31	33.72	27.42
Local scale	100.00	100.00	98.72	89.38	80.22
Sampling	100.00	100.00	100.00	100.00	99.23
Noise	100.00	100.00	100.00	100.00	100.00
Shot noise	100.00	100.00	100.00	100.00	100.00
Partial	7.54	5.70	4.51	3.58	2.95
Mixed	53.13	55.86	47.77	37.54	30.34
Average	94.94	93.12	90.84	87.82	85.00

Table 3: Performance (mAP in %) of ShapeGoogle using bags of features of size 48 based on SI-HKS local descriptor.

	Strength				
Transform.	**1**	**≤2**	**≤3**	**≤4**	**≤5**
Isometry	100.00	100.00	100.00	100.00	100.00
Topology	96.15	96.15	94.87	93.27	92.69
Holes	100.00	100.00	100.00	94.71	89.97
Micro holes	100.00	100.00	100.00	100.00	100.00
Scale	91.03	95.51	97.01	97.76	98.21
Local scale	100.00	100.00	97.44	89.38	82.08
Sampling	100.00	100.00	100.00	100.00	97.69
Noise	100.00	100.00	100.00	100.00	100.00
Shot noise	100.00	100.00	100.00	100.00	100.00
Partial	17.43	10.31	9.57	8.06	6.61
Mixed	56.47	57.44	63.59	67.47	65.07
Average	97.05	95.16	94.03	92.54	90.79

Figure 11 shows a few examples of retrieved shapes, ordered by relevance, which is inversely proportional to the distance from the query shape. Using HKS, all the matches for *scale* and *mixed* transformations queries (rows $2-3$ and 4) are incorrect (middle column). On the other hand, using the SI-HKS the results are mostly correct (right column).

7 Conclusion

In this paper we have introduced a method to construct scale invariant descriptors without relying on scale selection. Our experimental results demonstrate that these descriptors compare favorably to current state-of-the-art alternatives when evaluated on standard datasets. Moreover, by virtue of being independent of scale selection, our descriptors can be computed densely over

the signal domain.

In future work we intend to explore several avenues to expand their usefulness. In specific, we intend to pursue the use of scale-invariant image descriptors for object recognition, both for Bag-of-Words classifiers and for part-based models. We also intend to pursue the integration of Self-Similarity Descriptors [70] with our approach, as well as the construction of scale-invariant surface descriptors from open, noisy surfaces, such as those delivered by depth sensors.

Appendix

Denoting by $I_\alpha[n]$ the signal obtained by sampling $I(x/\alpha)$ as in (6), we prove that $I_\alpha[n] = I_1[n-1], \forall n$:

$$\begin{aligned} I_\alpha[n] &= \mathcal{I}_\alpha(c_0\alpha^n, c\alpha^n) = I_\alpha(x) * g_{c\alpha^n}(x)|_{x=c_0\alpha^n} \\ &= \int_t I_\alpha(t - c_0\alpha^n) g_{c\alpha^n}(t) dt \\ &= \int_t I_1(t/\alpha - c_0\alpha^{n-1}) \alpha g_{c\alpha^{n-1}}(t/\alpha) dt \\ &\overset{t'=t/\alpha}{=} \int_{t'} I_1(t' - c_0\alpha^{n-1}) g_{c\alpha^{n-1}}(t') dt' \\ &= I_1[n-1] \end{aligned} \tag{23}$$

By recursion we have that $I_{\alpha^k}[n] = I_1[n-k], \forall n, k$.

References

[1] J. Aflalo, A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Deformable shape retrieval by learning diffusion kernels. In *Proc. Scale Space and Variational Methods (SSVM)*, 2011.

[2] M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature-a quantum mechanical approach to shape analyis. In *Proc. CVPR*, 2011.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. J. Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[4] M. Ben-Chen, O. Weber, and C. Gotsman. Characterizing shape using conformal factors. In *Proc. Eurographics Workshop on Shape Retrieval*, 2008.

[5] A. Berg and J. Malik. Geometric blur for template matching. In *Proc. CVPR*, 2001.

[6] A. M. Bronstein. Spectral descriptors for deformable shapes. arXiv 1110.5015, 2011.

[7] A. M. Bronstein, M. M. Bronstein, A. M. Bruckstein, and R. Kimmel. Analysis of two-dimensional non-rigid shapes. *IJCV*, 78(1):67–88, 2008.

[8] A. M. Bronstein, M. M. Bronstein, U. Castellani, B. Falcidieno, A. Fusiello, A. Godil, L. J. Guibas, I. Kokkinos, Z. Lian, M. Ovsjanikov, et al. SHREC 2010: robust large-scale shape retrieval benchmark. 2010.

[9] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *PNAS*, 103(5):1168–1172, 2006.

[10] A. M. Bronstein, M. M. Bronstein, R. Kimmel, M. Mahmoudi, and G. Sapiro. A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *IJCV*, 89(2–3):266–286, 2010.

[11] A. M. Bronstein, M .M. Bronstein, M. Ovsjanikov, and L. J. Guibas. Shape google: a computer vision approach to invariant shape retrieval. In *Proc. NORDIA*, 2009.

[12] M. M. Bronstein and A. M. Bronstein. Shape recognition with spectral distances. *Trans. PAMI*, 33(5):1065–1071, 2011.

[13] M. M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Proc. CVPR*, 2010.

[14] M. Brown, G. Hua, and S. Winder. Discriminative learning of local image descriptors. *Trans. PAMI*, 33(1):43–57, 2011.

[15] H. Cai, K. Mikolajczyk, and J. Matas. Learning linear discriminant projections for dimensionality reduction of image descriptors. In *Proc. BMVC*, 2008.

[16] D. Casasent and D. Psaltis. Position, rotation, and scale invariant optical correlation. *Applied Optics*, 15(7):258–261, 1976.

[17] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *Proc. ICCV*, 2007.

[18] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21:5–30, July 2006.

[19] J. Digne, J. M. Morel, N. Audfray, and C. Mehdi-Souzani. The Level Set Tree on Meshes. In *Proc. 3DPVT*, 2010.

[20] G. Dorkó and C. Schmid. Maximally stable local description for scale selection. In *ECCV*, 2006.

[21] A. Elad and R. Kimmel. Bending invariant representations for surfaces. In *Proc. CVPR*, 2001.

[22] M. Felsberg and G. Sommer. The monogenic signal. *IEEE Trans. on Signal Processing*, 49(12):3136–3144, 2001.

[23] M. S. Floater and K. Hormann. Surface parameterization: a tutorial and survey. *Advances in Multiresolution for Geometric Modelling*, 1, 2005.

[24] F. Fraundorfer, H. Stewénius, and D. Nistér. A binning scheme for fast hard drive based image search. In *Proc. CVPR*, 2007.

[25] B. Fulkerson, A. Vedaldi, and S. Soatto. Localizing objects with smart dictionaries. In *Proc. ECCV*, 2008.

[26] K. Gebal, J.A. Bærentzen, H. Aanæs, and R. Larsen. Shape analysis using the auto diffusion function. In *Computer Graphics Forum*, volume 28, pages 1405–1413, 2009.

[27] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann. Robust global registration. In *Proc. SGP*, 2005.

[28] J. M. Geusebroek, A. W. M. Smeulders, and J. van de Weijer. Fast anisotropic gauss filtering. 12(8):938–943, 2003.

[29] G. Hua, M. Brown, and S. Winder. Discriminant embedding for local image descriptors. In *Proc. ICCV*, 2007.

[30] Q. X. Huang, S. Flöry, N. Gelfand, M. Hofer, and H. Pottmann. Reassembling fractured objects by geometric matching. *Trans. Graphics*, 25(3):569–578, 2006.

[31] H. Jégou, H. Harzallah, and C. Schmid. A contextual dissimilarity measure for accurate and efficient image search. In *Proc. CVPR*, 2007.

[32] T. Kadir and M. Brady. Saliency, scale and image description. *IJCV*, 45(2):83–105, 2001.

[33] A. Kannan, N. Jojic, and B.J. Frey. Fast transformation-invariant factor analysis. In *Proc. NIPS*, 2002.

[34] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Proc. SGP*, 2003.

[35] Y. Ke and R. Sukthankar. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In *Proc. CVPR*, 2004.

[36] I. Kokkinos and A. Yuille. Scale Invariance without Scale Selection. In *Proc. CVPR*, 2008.

[37] M. Kolomenkin, I. Shimshoni, and A. Tal. On edge detection on surfaces,. In *Proc. CVPR*, 2009.

[38] A. Kovnatsky, M. M. Bronstein, A. M. Bronstein, and R. Kimmel. Photometric heat kernel signatures. In *Proc. Conf. on Scale Space and Variational Methods in Computer Vision (SSVM)*, 2011.

[39] B. Lévy. Laplace-Beltrami eigenfunctions towards an algorithm that "understands" geometry. In *Proc. Shape Modeling and Applications*, 2006.

[40] T. Lindeberg. Feature Detection with Automatic Scale Selection. *IJCV*, 30(2):79–116, 1998.

[41] T. Lindeberg and L. Florack. Foveal scale-space and the linear increase of receptive feld size as a function of eccentricity. *CVIU*, 97(2):209241, 1996.

[42] H. Ling and D. Jacobs. Using the inner-distance for classification of articulated shapes. In *Proc. CVPR*, 2005.

[43] R. Litman, A. M. Bronstein, and M. M. Bronstein. Diffusion-geometric maximally stable component detection in deformable shapes. *Computers and Graphics*, 35(3), 2011.

[44] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 60(2):91–110, 2004.

[45] M. Mahmoudi and G. Sapiro. Three-dimensional point cloud recognition via distributions of geometric distances. *Graphical Models*, 71(1):22–31, 2009.

[46] S. Manay, B.W. Hong, A.J. Yezzi, and S. Soatto. Integral invariant signatures. In *Proc. ECCV*, 2004.

[47] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[48] F. Mémoli and G. Sapiro. A theoretical and computational framework for isometry invariant recognition of point cloud data. *Foundations of Computational Mathematics*, 5:313–346, 2005.

[49] M. Meyer, M. Desbrun, P. Schroder, and A. H. Barr. Discrete differential-geometry operators for triangulated 2-manifolds. *Visualization and Mathematics III*, pages 35–57, 2003.

[50] K. Mikolajczyk and C. Schmid. Scale and Affine Invariant Interest Point Detectors. *IJCV*, 60(1), 2004.

[51] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *Trans. PAMI*, 2005.

[52] K. Mikolajczyk, A. Zisserman, and C. Schmid. Shape recognition with edge-based features. In *Proc. BMVC*, 2003.

[53] N. J. Mitra, L. J. Guibas, J. Giesen, and M. Pauly. Probabilistic fingerprints for shapes. In *Proc. SGP*, 2006.

[54] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *Proc. ECCV*, 2006.

[55] A. Oppenheim, R. Schafer, and J. Buck. *Discrete-Time Signal Processing*. Prentice-Hall, 1999.

[56] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Trans. on Graphics (TOG)*, 21(4):807–832, 2002.

[57] M. Ovsjanikov, A. M. Bronstein, M.M. Bronstein, and L .J. Guibas. Shape Google: a computer vision approach to invariant shape retrieval. In *Proc. NORDIA*, 2009.

[58] M. Ovsjanikov, J. Sun, and L. Guibas. Global intrinsic symmetries of shapes. *Computer Graphics Forum*, 27(5):1341–1348, 2008.

[59] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proc. CVPR*, 2007.

[60] J. Philbin, M. Isard, J. Sivic, and A. Zisserman. Descriptor learning for efficient retrieval. In *Proc. ECCV*, 2010.

[61] U. Pinkall and K. Polthier. Computing discrete minimal surfaces and their conjugates. *Experimental mathematics*, 2(1):15–36, 1993.

[62] M. Porat and Y. Zeevi. The generalized gabor scheme of image representation in biological and machine vision. *Trans. PAMI*, 10(4):452–468, 1988.

[63] D. Raviv, A. M. Bronstein, M. M. Bronstein, R. Kimmel, and N. Sochen. Affine-invariant diffusion geometry for the analysis of deformable 3d shapes. In *Proc. CVPR*, 2011.

[64] D. Raviv, M. M. Bronstein, A. M. Bronstein, and R. Kimmel. Volumetric heat kernel signatures. In *Proc. ACM Multimedia Workshop on 3D Object Retrieval*, 2010.

[65] M. Reuter, S. Biasotti, D. Giorgi, G. Patanè, and M. Spagnuolo. Discrete Laplace–Beltrami operators for shape analysis and segmentation. *Computers & Graphics*, 33(3):381–390, 2009.

[66] M. Reuter, F.-E. Wolter, and N. Peinecke. Laplace-spectra as fingerprints for shape matching. In *Proc. ACM Symp. Solid and Physical Modeling*, pages 101–106, 2005.

[67] R. M. Rustamov. Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In *Proc. SGP*, 2007.

[68] S. Belongie and J. Malik and J. Puzicha. Shape matching and object recognition using shape contexts. *Trans. PAMI*, 24(4):509–522, 2002.

[69] E. L. Schwartz. Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25(4):181–194, 1977.

[70] Eli Shechtman and Michal Irani. Matching local self-similarities across images and videos. In *Proc. CVPR*, 2007.

[71] I. Sipiran and B. Bustos. A robust 3D interest points detector based on Harris operator. In *Proc. 3DOR*, 2010.

[72] J. Sivic and A. Zisserman. Video Google: a text retrieval approach to object matching in videos. In *Proc. CVPR*, 2003.

[73] C. Strecha, A. M. Bronstein, M. M. Bronstein, and P. Fua. LDAHash: Improved matching with smaller descriptors. *Trans. PAMI*, 34(1):66–78, 2012.

[74] J. Sun, M. Ovsjanikov, and L. Guibas. A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion. In *Computer Graphics Forum*, volume 28, pages 1383–1392, 2009.

[75] A. Tal, M. Elad, and S. Ar. Content based retrieval of VRML objects - an iterative and interactive approach. In *Proc. Eurographics Workshop on Multimedia*, 2001.

[76] N. Thorstensen and R. Keriven. Non-rigid shape matching using geometry and photometry. In *Proc. CVPR*, 2009.

[77] Engin Tola, Vincent Lepetit, and Pascal Fua. A fast local descriptor for dense matching. In *Proc. CVPR*, 2008.

[78] R. Toldo, U. Castellani, and A. Fusiello. Visual vocabulary signature for 3D object retrieval and partial matching. In *Proc. 3DOR*, 2009.

[79] M. Wardetzky, S. Mathur, F. Kälberer, and E. Grinspun. Discrete Laplace operators: no free lunch. In *Conf. Computer Graphics and Interactive Techniques*, 2008.

[80] S. Winder, G. Hua, and M. Brown. Picking the best daisy. In *Proc. CVPR*, 2009.

[81] G. Wolberg and S. Zokai. Robust image registration using log-polar transform. In *Proc. ICIP*, 2000.

[82] A. Zaharescu, E. Boyer, K. Varanasi, and R Horaud. Surface feature detection and description with applications to mesh matching. In *Proc. CVPR*, 2009.

[83] C. Zhang and T. Chen. Efficient feature extraction for 2D/3D objects in meshrepresentation. In *Proc. ICIP*, volume 3, 2001.

Contents

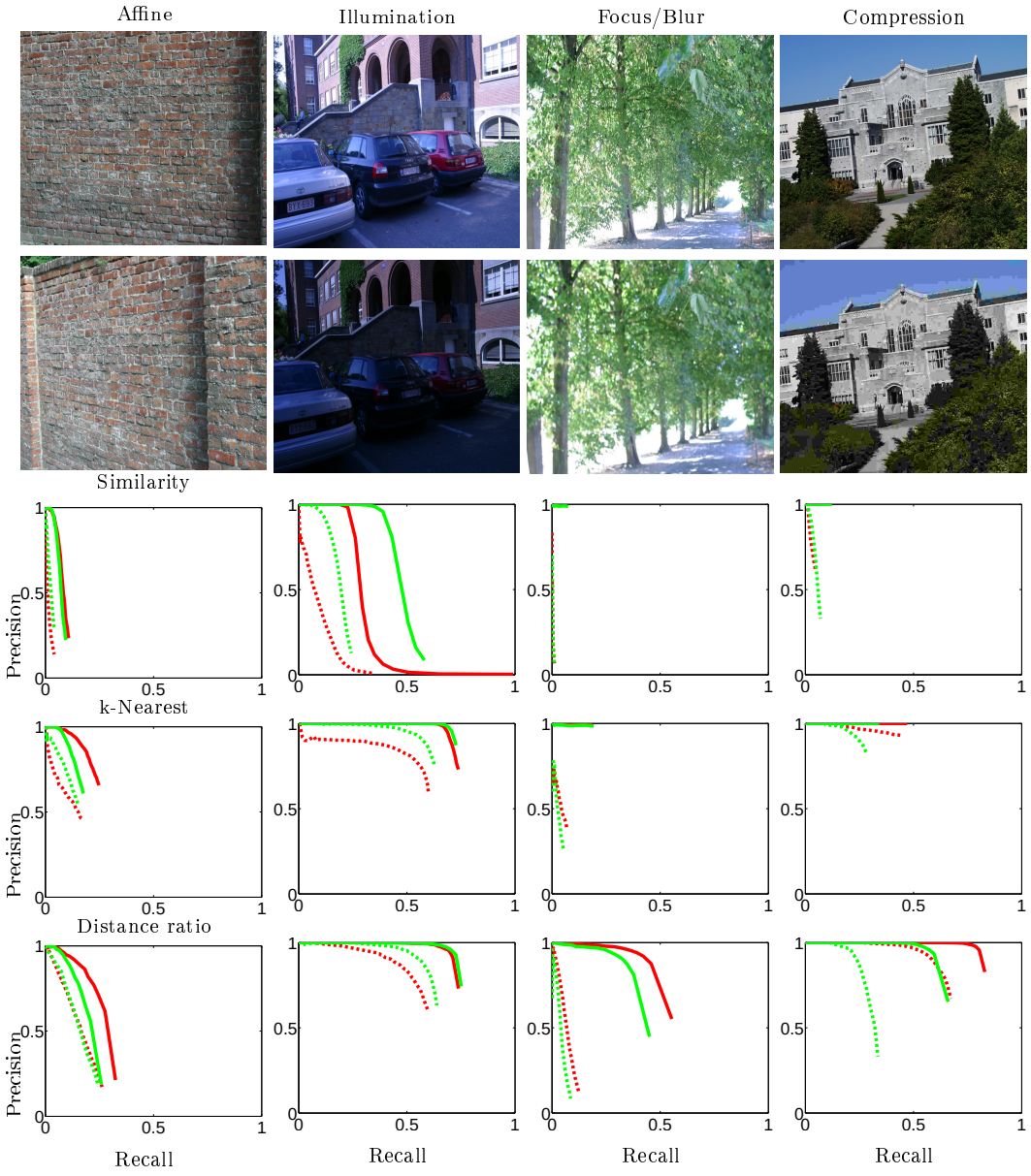


Figure 8: Precision-Recall curves for transformations other than scale and rotation: we compare SIFT (dashed) to our descriptors (solid) on Harris-Laplace (green) and Hessian-Laplace interest points (red).

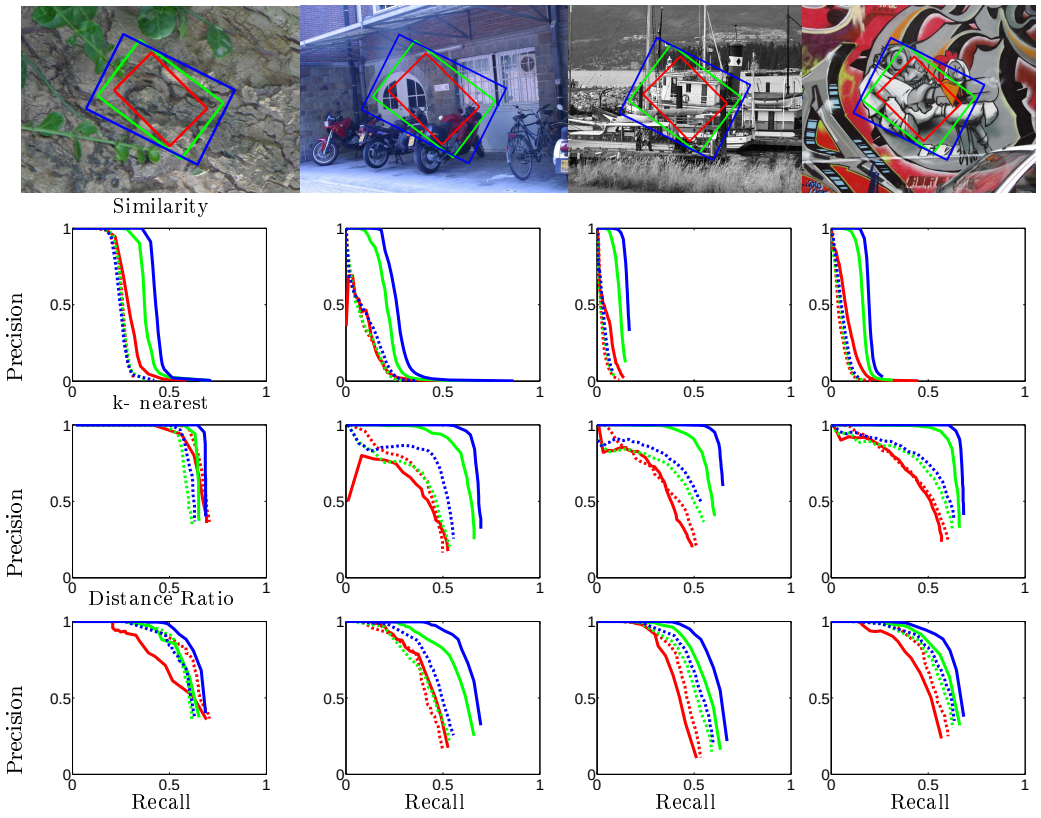


Figure 9: Precision-Recall curves comparing SIFT (dashed) to our descriptor (solid) for synthetic scaling transformations. We use a range of simulated scale- σ and rotation-θ changes, visualized with differently colored boxes and curves; red corresponds to $(\sigma = .27, \theta = \pi/4)$, green to $(\sigma = 0.33, \theta = \pi/5)$, and blue to $(\sigma = 0.42, \theta = \pi/8)$.

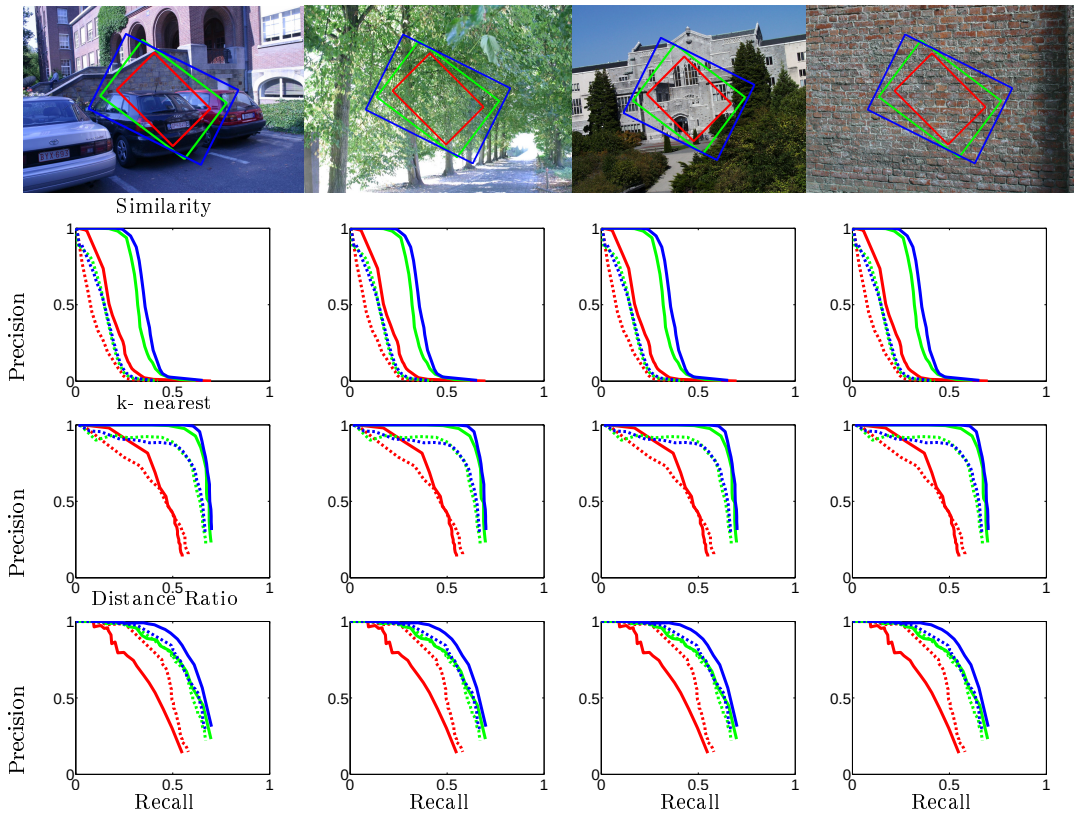


Figure 10: Precision-Recall curves comparing SIFT (dashed) to our descriptor (solid) for synthetic scaling transformations. We use a range of simulated scale- σ and rotation-θ changes, visualized with differently colored boxes and curves; red corresponds to $(\sigma = .27, \theta = \pi/4)$, green to $(\sigma = 0.33, \theta = \pi/5)$, and blue to $(\sigma = 0.42, \theta = \pi/8)$.

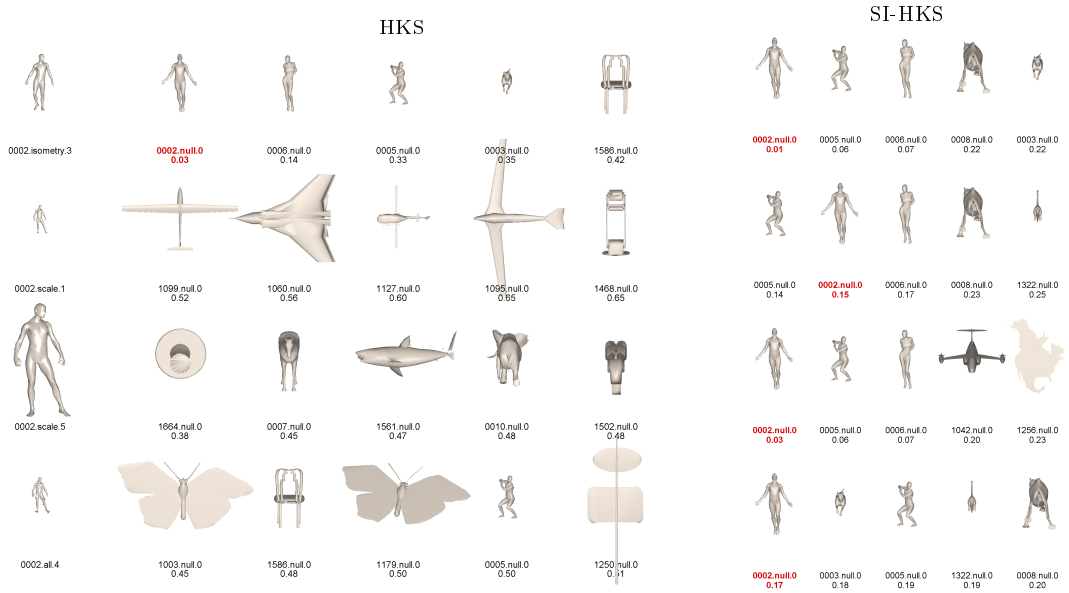


Figure 11: Examples of first five matches for queries (column 1) found using bags of features based on HKS (columns 2-6) and SI-HKS (columns 7-11). Correct matches are shown in red. Only one match is correct, and ideally it should be the first.

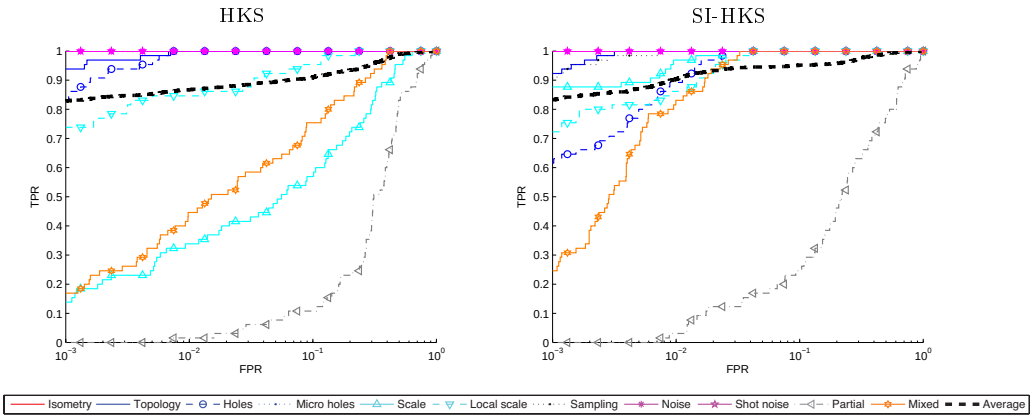
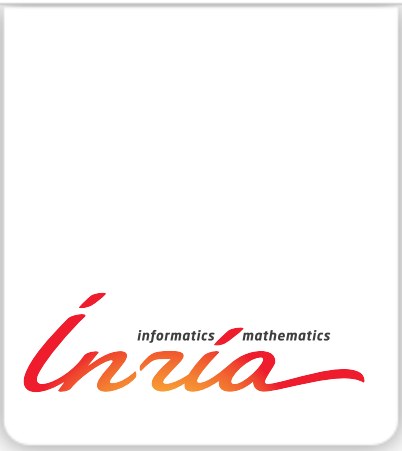


Figure 12: ROC curves showing shape retrieval performance using bags of features based on HKS and SI-HKS.



RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

Parc Orsay Université
4 rue Jacques Monod
91893 Orsay Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399